ELSEVIER

Mechanisms of Development 110 (2002) 71–83

www.elsevier.com/locate/modo

A P-insertion screen identifying novel X-linked essential genes in *Drosophila*

Henri-Marc Bourbon[a,1], Geneviève Gonzy-Treboul[b,1], Frédérique Peronnet[c,1], Marie-Francoise Alin[b], Claude Ardourel[a], Corinne Benassayag[a], David Cribbs[a], Jean Deutsch[c], Pierre Ferrer[a], Marc Haenlin[a], Jean-Antoine Lepesant[b], Stéphane Noselli[a,2], Alain Vincent[a,*]

[a]*Centre de Biologie du Développement, UMR 5547 CNRS/UPS, 118 Route de Narbonne, 31062 Toulouse Cedex, France*
[b]*Institut Jacques Monod, UMR 7592 CNRS/Universités Paris 6 et Paris 7, 2 Place Jusieu, 75251 Paris Cedex 05, France*
[c]*Développement et Evolution UMR 7622 CNRS/Université Paris 6, 7 Quai St Bernard, 75252 Paris Cedex 05, France*

Received 7 June 2001; received in revised form 5 September 2001; accepted 7 September 2001

Abstract

The recent determination and annotation of the entire euchromatic sequence of the *Drosophila melanogaster* genome predicted the existence of about 13 600 different genes (Science 287 (2000) 2185; http://www.fruitfly.org/annot/index.html). In parallel, the Berkeley Drosophila Genome Project (BDGP) has undertaken systematic P-insertion screens, to isolate new lethals and misexpressing lines. To date, however, the genes of the X chromosome have been under-represented in the screens performed. In order both to characterize several X-linked genes of prime interest to our laboratories and contribute to the collection of lethal P-insertions available to the community, we performed a P-insertion mutagenesis of the X chromosome. Using the PlacW and PGawB P-elements as mutagens, we generated two complementary sets of enhancer-trap lines, $l(1)_TPL$ and $l(1)_TPG$, respectively, which both contain a reporter gene whose developmental expression can be monitored when driven by nearby enhancer sequences. We report here the characterization of 260 new insertions, mapping to 133 different genes or predicted CGs. Of these, 83 correspond to genes for which no lethal mutation had yet been reported. For 64 of those, we could confirm that lethality was solely due to the P-element insertion. The primary molecular data, reporter gene expression patterns (observed in embryos, third instar larvae and adult ovaries) and proposed CG assignment for each strain can be accessed and updated on our website at the following address: http://www-cbd.ups-tlse.fr:8080/screen.

Keywords: Euchromatic sequence; *Drosophila melanogaster*; P-insertion screen

1. Introduction

The exploding field of genome sequencing identifies new genes at an unprecedented pace, as compared to other methods. The recent determination of the entire euchromatic sequence of the *Drosophila melanogaster* genome predicted the existence of about 13 600 different genes (Adams et al., 2000), including an estimated 3600 vital loci (Miklos and Rubin, 1996). By comparison, only about 2500 different genes, representing the outcome of a century of genetic analyses, were molecularly characterized by the research community prior to the genome sequencing project. Classification of predicted *Drosophila* proteins into large families, according to conservation of structural motifs (based on Interpro: http://www.ebi.ac.uk/interpro/) has been included in the annotated *Drosophila* genome (Rubin et al., 2000). With the availability of loss-of-function and gain-of-function approaches to study developmental gene function (Rorth et al., 1998), the annotated genome sequence of *D. melanogaster* has become a sophisticated tool for functional and evolutionary comparative studies (Rubin and Lewis, 2000). P-element strains that each mutate a unique open reading frame (ORF) already collected from the Berkeley Drosophila Genome Project (BDGP; http://www.fruitfly.org/; Spradling et al., 1999), and previous mutageneses carried out in several laboratories, account for about 1200 single-gene disruptions. Yet, these insertional mutageneses have mainly targeted the autosomes, such that the genes of the X chromosome (about one-sixth of the total gene number) are on the whole under-represented in the current P-lethal insertion collections. Our laboratories became interested in X-linked genes for which

* Corresponding author. Tel.: +33-5-6155-8289; fax: +33-5-6155-6507.
E-mail address: vincent@cict.fr (A. Vincent).

[1] The first three authors contributed equally to the work.

[2] Present address: Institut de Recherches 'Signaling, Developmental Biology and Cancer' UMR 6543 – CNRS/University de Nice, Parc Valrose 06108 Nice Cedex 2, France.

0925-4773/02/$ - see front matter
PII: S0925-4773(01)00566-4

no tagged mutations were available (Gonzy-Treboul et al., 1995; Glise et al., 1995; Mevel-Ninio et al., 1995). In order to genetically characterize these genes, identify novel genes functionally related to them, and contribute to the collection of lethal P-insertions available to the community, we undertook a P-insertion mutagenesis of the X chromosome. The employed mutagenic transposons, PlacW (Bier et al., 1989) and PGawB (Brand and Perrimon, 1993), both contain a reporter gene whose developmental expression can be driven by nearby enhancer sequences. Taking advantage of the annotated genome sequence (http://www.fruitfly.org/annot/index.html), we localized each of 260 insertions generated in our screen on the chromosome by identifying P-adjacent sequences. We report here their association with known genes or predicted CGs on the basis of genomic sequence and expression data (ESTs). Among 133 different targeted genes (as defined by a CG number), 83 represented genes for which no lethal mutation was known. P-element excision experiments confirmed that the P-insertion was directly responsible for the lethality in 64 out of 80 cases tested, revealing many new X-linked essential genes, and suggesting that by and large, at least 80% of the mutant strains isolated in our screen are linked to the mapped P-insertion. When possible, we indicate the predicted function of novel essential genes identified in our screen, on the basis of their structure and/or previous classification by Goldstein and Gunawardena (2000), Lasko (2000) and Mount and Salz (2000).

To gain further insight into the potential developmental roles of genes represented in our $l(1)_T$ collection, the reporter gene expression pattern was examined in embryos and ovaries for all mutant strains. Third instar larval tissues were also examined in most strains. These expression patterns are summarized in a table using a simple keyword nomenclature and made available for browsing at the following internet address: http://www-cbd.ups-tlse.fr:8080/screen.

2. Results

2.1. Establishment of a collection of lethal P-element insertions on the X-chromosome

A total of 279 independent insertions on the X-chromosome were characterized, as described in Fig. 1. For the sake of specificity and versatility, two different enhancer-trap starting elements were used, PlacW and PGawB. A minority of the ~45 000 individual crosses set up for Cross 1 were unproductive (16% for PL and 6% for PG lines). Of the remaining dysgenic crosses, the selection Cross 2 led to the recovery of 110 $l(1)_T$PL (out of 4660 crosses) and 169 $l(1)_T$PG lethals (out of 7450 crosses). Thus, similar frequencies of one lethal PL insertion per 42 female candidates and one lethal PG insertion per 44 female candidates were observed. A strong bias against the identification of viable X-linked mutations displaying morphological defects was present in this screen, since the Cross 2 progeny were initially screened visually in their vials. Among the putative lethal lines isolated, several were subsequently lost because of poor female fertility. The $l(1)_T$ collection currently contains 260 independent lines.

2.2. Characterizing insertions using flanking DNA sequences

The total number of different known or predicted (CG) genes for which we obtained at least one insertion is 133. We recovered a single insertion in 75 CGs (32 PL and 43 PG); two independent insertions in 33 CGs; three in 13; and from four to 13 (in the *inx2* locus) in 12 others (Table 1 and Fig. 1A). Based on the non-recovery of viable males following re-mobilization of the P-element, we suspect that lethality in a small proportion of mutant lines may not be due to the P-insertion but due to mutations elsewhere on the chromosome. Among the 133 loci (P insertion/CG) represented in our $l(1)_T$ collection, 83 of them (60%) correspond to CGs for which there were no known pre-existing mutations (GadFly releases 1 and 2, October 2000). This last number may nevertheless be overestimated, due to the average delay between characterization of new mutants and their publication. While in most cases the link between a P-insertion position and a gene appeared straightforward based on available genome annotation, this was not true for all $l(1)_T$ strains. In difficult cases, such as the genes *perlecan*, and *CG9650*, for which we obtained several independent insertions, we undertook a detailed reappraisal of the genomic region and available EST information. This led us to modify the previously proposed structure of these genes (Fig. 1B). Similarly, we noticed that the putative *Drosophila* orthologs of midnolin ($l(1)_T$PL77) and lecithin–retinol acetyltransferase ($l(1)_T$)PG19, 41, 102) were each previously described as two separate CGs, CG9725 + 9732 and CG11595 + 12482, respectively (Table 1). Finally, in the following cases ($l(1)_T$ PL25, 49, 60#1, 81, 109, PG21, 76, 80, 112, 133, 150, 166), we were unable to identify with confidence the mutated CG, solely based on the genome annotation. The nearest CG and putative ORF to insertions $l(1)_T$ PL30, 64, 65#1 and PG42, 99, 136, 157, 164 is *rutabaga* (*rut*). Based on existing data, however, (Levin et al., 1992), we suspect that these insertions do not affect *rut* but rather another gene which remains to be identified.

2.3. Embryonic lethals displaying cuticular defects

Very few systematic large-scale mutagenesis screens have been conducted on the X-chromosome. In their pioneering EMS screen for mutations affecting the pattern of the larval cuticle, Wieschaus et al. (1984) reported that about 20% of lethal point mutations caused death during embryogenesis; of these, only one-sixth, or 3%, of the total lethals were associated with visible defects in the final cuticle pattern. About 20 of those loci correspond to key developmental genes, most of which have now been

characterized at the molecular level. From our P-element screen, we recovered mutations in about half of these patterning genes, including *armadillo* (*arm*) (Fig. 2A), *cut* (*ct*), *giant* (*gt*), *hindsight* (*hnt*), *lethal-myospheroid* (*mys*), *Notch* (*N*), *shavenbaby* (*svb*) and *short gastrulation* (*sog*). We also recovered three independent mutations in *brinker*, which encodes a transcriptional repressor involved in regulating the transcriptional response to the morphogen Decapentaplegic (Jazwinska et al., 1999). Few other loci caused recognizable cuticular defects when mutated. Mutations in CG1521, which encodes a potential membrane-associated divalent cation transporter, result in embryonic death before the cuticle is completely deposited with melanotic depots specifically found in the salivary glands and dorsal branches of the tracheae (Fig. 2A). A mutation in CG12235 (*Arp11*), which encodes an actin-binding component of the Arp2/3 protein complex, results in a poorly differentiated cuticle. Mutations in CG12701, encoding a C_2H_2 zinc finger protein, show specific head defects.

2.4. New mutants in genes encoding nuclear components

Among those genes for which no mutants have been previously reported, we recovered insertions in various types of transcription factors. These include C_2H_2 zinc finger proteins (CG9650, 12701), the Myc/Mad/Mnt related b-HLH protein D-Mnt (CG2856; Peyrefitte et al., 2001), a new Rel-domain containing protein (CG11172), an orphan nuclear receptor related to Ftz-F1 (CG16902), and Smr, a co-repressor of the Ecdysone receptor. We also recovered mutations in genes encoding a putative nuclear transport factor related to the human placental protein NTF-2

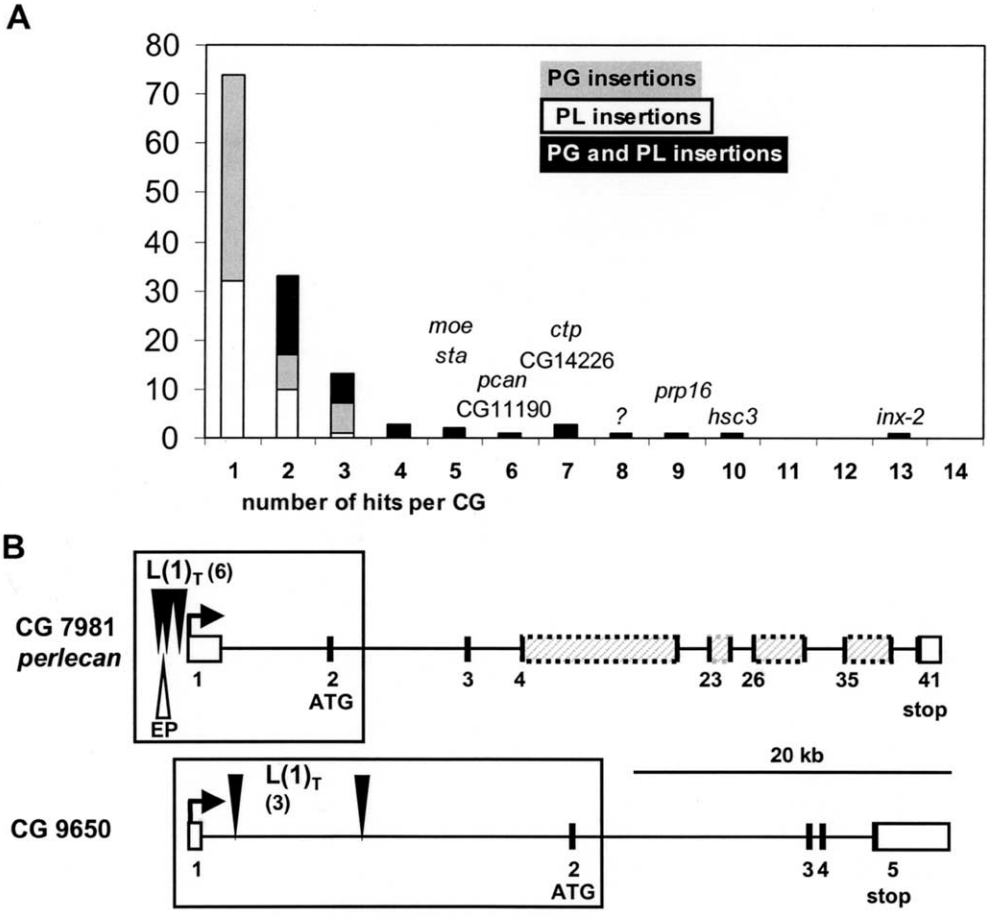

Fig. 1. (A) Frequency distribution of the 260 single P-element lethal insertions on the X-chromosome generated in our screen. The genes or CGs for which we recovered five or more independent insertions are indicated. (B) Revised exon–intron structures of the *perlecan*/*CG7981* and *CG9650* loci indicating the positions of the $l(1)_T$ insertions (black arrowheads). The white arrowhead indicates an EP insertion (EP1160) from the BDGP collection. Filled boxes represent the protein-coding regions, based on the positions of predicted translational initiator and stop codons. Exon numbers are indicated. The position of the first two, not previously recognized, exons of *pcan* were deduced both from the identification of two consensus donor splice sites (tgcaGTGAGT and ggtgGTAAGT) and a canonical translation initiator ATG, and the structure of ESTs GM10545, GM01493 and GM02428 whose 3′ ends correspond to an internal polyA sequence. Note that a putative signal peptide (MGRRLRAAFWLLAVLIVIE) is predicted, when using the SignalP software (http://www.cbs.dtu.dk/services/SignalP). The revised structure of CG9650 was deduced from comparison of the genomic sequence and the available EST LD11946 whose 5′-untranslated sequences is split by an intron 24 617 bp long. Predicted introns not included in the annotated genome sequence release 2, are boxed.

Table 1
List of strains from the primary $l(1)_T$ collection of P-element insertions[a]

Cyto	l(1) T line	CG	Gene/predicted function	Sequence similarities
1B5–6	PG159	4262	*elav; RNA binding factor*	
1B7	PG33, 47	4122	*svr* (*silver*); arginine carboxy peptidase	
1B10–11	PG49	16 983	*skpA* S-phase kinase	
1B13	PG29	7622	*rpl 36*; ribosomal protein	
1C1	PL62, 70	7434*	*rpl 22*; ribosomal protein	
1F2	*PL102*	*3638**	Transcription factor	C_2H_2 zinc fingers
2A2–3	PG118	3719*	Thioredoxin-like	
2A2–4	PG97*	14 772*	Motor protein?	
2B1–2	PL33, (65 #2), 79, 93, PG11	14–792	*sta* (*stubarista*); ribosomal component	
2B5	**PL86***, **PG12**, **83**	11 491	*br-C* (*broad-Complex*); transcription factor	
2B13–16	PG9*	3587*		kti12; ATP/GTP binding motif
2B15	PL27, PG98,120	11 579	*arm* (*armadillo*); beta-catenin	
2C2	PL78*	16 902*	Steroid hormone nuclear receptor	Ftz-F1
2C5–6	PL1*, PG13	4281*		Xenopus Nfr1
2C8–9	PG95,153*	3981*	Synaptogamin interacting protein	Nematode Unc-76; human zygin
3A2	PL20	7952	*gt* (*giant*); transcription factor	
3A4	PL63*, PG68, 93, 94, 122*, 123	7981*	*pcan* (*perlecan*)	Heparan sulfate proteoglycan
3A4	PL109, PG150, 166	12 497/13 758		LDL receptor/G-protein coupled receptor
3B2–3	PL36	2621	*sgg* (*shaggy*); protein kinase	
3C7–9	PL24, PG50	3936	*Notch*; transmembrane receptor	
3D5	PL35*, *51*, PG45*	10 798	*dm*, *D-myc*; b-HLH transcription factor	
3F3	*PG114*	2856*	*D-mnt*; b-HLH transcription factor	Mouse Mnt: Max-interacting protein
3F3	**PL56#1**, **PG69**, **89**	2849*	*Ral-a*; Ras related protein; cell cycle regulator	
3F4	PL56 #2	12 462*	Kinase	Tousled
4B2–3	PL38, 84*, 108	3665	*Fas II* (*fasciclinII*)	
4C5–6	PL67, PG37*	12 212	*pebled/hindsight* (*peb/hnd*) transcription factor	C_2H_2 zinc fingers
4D1	PL13, PG10, 54, 88, 134, 139, 145	6998	*ctp* (*cut-up*) dynein light chain	
4D1	PL28*#2	4068*/6899	Ptp 4 E (tyrosine protein phosphatase)	
4E1	**PL107***	6824	*svb* (*shavenbaby:ovo*) transcription factor	C_2H_2 zinc fingers
4F9	*PG126*	15 784*	Domain kininogen?	
5B2–5	*PL99*, PG135*	3125*	Putative tumor suppressor	EGF type repeats
5C2–5	PG28*, PG40*	4027	*Act 5C*	
5D6	*PL59*	3599*	GPI-linked vanin	
5D6	*PL76*	3774*	UDP-galactose transporter	
5D6	PL91*	3595	*sqh* (*spaghetti-squash*); type II myosin	
5E1	PL83*	12 225*	*spt 6*; chromatin associated protein	
5E5–6	PG129*	3861*	Citrate synthase mitochondriale	
6E4	PG27, 147*	14 430*	SH3 domain binding protein	
6E4–5	**PL19**, **PG4**, **46**, **55**, **57**, **71**, **73**, **81**, **92**, **104***, **109***, **144**, **160**	4590*	*inx 2* (*prp 33*); innexin	
7A1–2	PG64*, 75*, 105	9650*	Transcription factor	C_2H_2 zinc fingers
7A3	PL49*	2059/1677*		
7A4	PG18, 91, 121	9653	*brk* (*brinker*) transcription factor	
7A4	PG155*	2079*	Ras GAP binding protein	
7A4	PL55#2*	1435*	CBP calcium binding protein	
7B3–5	PG142*	11 387	*cut*; transcription factor	Homeodomain protein
7C6–8	PL17* (*viable*)	*1422**	*transcytosys associated protein*	*P115 vesicular transport factor*
7C8	PG127*	10 777*	RNA metabolism	DEAD box protein
7DI	PL4, 53*	2206*		
7D I	PL81*	1531*/2206	E3 ubiquitin protein ligase	
7D1	PL37, 85*	1530*	E3 ubiquitin protein ligase	
7D6–8	PL31, PG79	1560	*mys* (*myospheroid*) intregin beta-subunit	
7D18	PL98*, PG17	2151*	*Gr* Glutathione reductase enzyme	
7D8–9	PG165*	2257*	Ubiquitin conjugating enzyme	
7E2–3	PL26, 50*, 71*, 60#2*, PG62, 111	11 190*		Orthologs in human, worm
7E5	PL28#1	18 009–11 195	*Trf2*, transcription initiation factor	TBP-like protein
7F1	PG80*, 112*	1521*/ 11 219*	Unknown/pip82	

Table 1 (*continued*)

Cyto	l(1) T line	CG	Gene/predicted function	Sequence similarities
8B4–7	**PL54***, **106***, **PG 26***, (**82**), **67***	10 701*	*moe* (*moesin*); cytoskeletal protein	
8C11	PL104*	15 365*		Leucine zipper human protein
8C16	PG162	12 664	*ld14*	
8E7–10	*PL52*	2994*	Membrane surface glycoprotein	Signal sequence receptor subunit
8E10–12	PG53*, 149, 167	3001*	Hexokinase II	
9B14–15	PL94*	17 841*		
9E2–3	PL45, 66, PG87	11 485	*ras* (*rasberry*)	
9E3–4	PL77*	9725–9732*	Nucleolar component	Midnolin
9E7–8	PG106, 141, 154	16 944	*sesB*; mitochondrial transport protein	
10B2–3	PG 1*	1938*	Chighen dynein light chain	
10B11	PL32, PG100	1725	*dlg 1* (*discs-large*); guanylate kinase	
10B15–C2	*PG107*	15 194*		
10C2	PL22, 39, PG124		*RpII215*; RNA polymerase subunit	
10C3	PG23*	1751*	Microsomal signal peptidase	
10C6–7	PL16*	1703*	ABC transporter	
10D8	PG70*	2446*		Arabidopsis protein
10E3–4	PL11*, 68, 73, PG 6, 34, 39, 65, 84, 86, 119	4147*	*hsc 3* (heat-shock cognate protein 3)	Adenosine triphosphatase; ER chaperone
10F4	PG48*	2025*	N-arginin convertase	
11B6–7	*PG22*, (32)	15 926*		
11B13	PG44*	2543*	Folylpolyglutamate synthetase-like	
11B16–17	**PL6***	12 719*	*Smr*; transcription factor	
11B17–18	PG133	18 137/3989		
11D6	*PL61*, 69*	12 714*		
11D9–10	PL57, PG132	1903	*sno*; nuclear protein	
11E6–9	PG30	1771	*mew*; integrin alpha-subunit	
11E11	PG151	1660		
12A4–B1	PG58, 128*	11 172*	*NFAT*; transcription factor	Rel-domain containing protein
12B2–4	*PL90*	11 428*		
12B8–9	*PG148*	10 992*	Endopeptidase; cathepsin B-like	
12DI–2	PL25*	18 319*/ un.ORF	*ben*/mitochondrial protein	
12D4	PL23*, 82*, 97, PG2, 15, 60*, 90, 108, 113	1405*	*D-prp16*; RNA splicing factor	Human Prp16
12D5	PG19*, 41, 102	11 595*–12 482*	Lecithin–retinol acetyltransferase	
12E10	*PL5*	9413*	Amino-acid transporter; permease	
12F7	PL30, 64, 65#1, PG42, 99, 136, 157, 164		*Nearest CG: rut*; Ca^{2+}/Cam adenylate cyclase	
13A8	PL95*, 105	5530*		
13A11–B1	PG138*	5599*	Mitochondrial acyltransferase	
13C5–8	PG63	9151	*acj 6* (*abnormal chemical jump*) transcription factor	Homeodomain protein
13D	PL7, 74, PG7, 16	9224	*sog* (*short gastrulation*)	
13F1	PL44	8497	*Rhp* (*Rhophilin*); Rho binding protein	PDZ domain protein
13F2	PG59	9012	*chc* (*chlathrin heavy chain*)	
14A6	*PG61*	9214*	Tumor suppressor	Tob
14B12	PL48, 103	3525	*eas* (*easily shocked*) (ethanolamine kinase)	
14B9–12	PG76	3415/9910	Katanin-80/estradiol 17 beta dehydrogenase like	
14D1	PL10, PG116	9907	*para*	
14E1	PL42*, PG156	9906*	Calnexin, chaperone	
14F1–2	PG169*	13 014*		
15A5–6	PL43*	9606*	Exosome protein	Human PMScl 75 autoantigen
15F1	*PL110*	12 996*		
16A1	PL9, PL21	5529	*Bar H1* transcription factor	
16B10	PG96, 131	5870	*beta-spectrin*; cytoskeletal protein	

Table 1 (*continued*)

Cyto	l(1) T line	CG	Gene/predicted function	Sequence similarities
16B4	PG51*	8465*		Ankyrin repeats
16D2	PG3*	6769*	Transcription factor	Yeast C_2H_2 zinc finger protein
16F3–7	PG115	7113	*scully* 3-hydroxylacyl-CoA dehydrogenase	
16F7	PL87*, PG31	7178	*wupA*/*tn1*; troponin 1	
17B5–C1	PL60#1*, PG137	15 047	*Bx* (*Beadex*); transcription factor	LIM domain protein
17C3–5	PL12*, PL55#1, PL72*, PG103	6606*	Cytokine receptor	Domain ATPase
17E4	*PL75*	7326*		
17E4–6	**PL29#2***	17 757*		
17E5–6	**PL29#1***	7307*		Human and *C. elegans* orthologs
18B1	*PG43*	7502*		
18B4–5	PL46*, PL47#1, PG161	12 359*	Sumo-protease	Sentrin
18B9–10	PL40*	7893*	*D-vav*	*vav* proto-oncogene
18C1–3	PG101*	8062*	Monocarboxylate transporter	
18C8–DI	PL3, PG85	12 202*	N-acetyl transferase	
18D1–3	PG140*	12 233*	Isocitrate dehydrogenase	
18D3–6	**PG36**, **66***	3917*	*D-grip 84*; microtubule-binding protein	
18D8–11	PG143*	12 238*	Transcription factor	Requiem
18D10–11	PL101*	12 235*	Actin-related protein	ARP11
18D13	**PL41*, 58*, 100*, PG5*, 14, 35*, 125**	14 226*	Tyrosine kinase receptor phosphatase	Myotactin
18E	*PL47#2*	14 216	Transcription initiation factor	
18F2–3	**PG8, PL92***	12 701	Transcription factor	C_2H_2 zinc finger protein
19A	PL15	11 937	*amn* (*amnesiac*)	
19E6	**PL18***, **88***	1740*	Nuclear transport factor, NTF2	NTF2
19F1–2	PG25*	1494*		Human gene linked to Stardgard's disease
19F5–6	PG117	14 619	Endopeptidase; ubiquitin-specific protease	
20BC	PG21	?	No ORF detected	
?	PG38*	17 131*	Membrane protein	SP71

[a] Strains from the primary $l(1)_T$ collection of lethal P-element insertions are listed according to their estimated physical order (cytology) along the first chromosome. The table gives the following information: cytogenetic location (based on the Gadfly annotation), number (PL*xx* or PG*xx*) of the mutant strain(s), CG number, disrupted gene (italics) and/or proposed biochemical function, and name(s) of known structural homologs in other eucaryotes. CGs for which no mutation has yet been reported in Flybase (http://flybase.bio.indiana.edu/.bin/fbidq.html) are indicated by an asterisk in the CG column. Underlined CG numbers indicate those for which we propose a revised structure (see main text and Fig.1B; bourbon@cict.fr). For a fraction of strains, we verified that recessive lethality was attributable to the P-insertion (see Section 3); these are indicated by asterisks in the $l(1)_T$ column. Conversely, this could not be verified for lines indicated in italics. The numbers in bold indicate mutant strains presently studied in our laboratories. $l(1)_T$ strain numbers in parentheses indicate insertions at the same nucleotide position than the preceding strain.

(CG1740), the *Drosophila* chromatin-associated protein SPT6 (CG12225), an exosome protein related to human PM-Scl-75 autoantigen (CG9606), and the *Drosophila* ortholog of the nucleolar protein midnolin (CG9725-9732).

2.5. New mutants in genes encoding cytoskeletal components

The *Drosophila* genome sequence has been systematically surveyed for its repertoire of different types of proteins, including cytoskeletal components (Goldstein and Gunawardena, 2000). Because of our specific interest in changes in epithelial cell morphology occurring during embryonic development (Glise et al, 1995; Vincent et al., 1997; Payre et al., 1999), we examined in detail the number and type of cytoskeletal components for which we obtained lethal insertions. Goldstein and Gunawardena (2000) listed 42 different CGs encoding putative cytoskeletal components. Our screen produced mutations in 11 of these, among them one new mutant each in the genes encoding the Dynein light chain 2 (*Dl1C2*, CG1938), Actin related protein 11 (*Arp11*, CG12235), moesin (*moe*, CG10701), microbule-associated protein Dgrip84 (CG3917) and possibly kataninP80 homolog (CG9910). We also recovered mutants in additional putative cytoskeletal components or cell-adhesion molecules corresponding to CG3599 (GPI-linked vanin) and CG8497 (Rhophilin, a PDZ domain Rho binding protein).

2.6. Patterns of reporter gene expression

Many of the lethal P-insertions were located immediately upstream of predicted transcription units or within mRNA 5′UTRs. About half of these mutations were associated with an expression of the reporter gene pattern in ovaries, embryos, third instar larval tissues, or a combination. To

describe these expression patterns, we adopted the simplified nomenclature described in the legend of Table 2. All corresponding pictures are displayed in our website. Fig. 2 shows selected patterns. The $l(1)_T$ PL105 insertion maps close to CG5530, a gene previously identified as a target of *twist* in the mesoderm of early embryos (Casal and Leptin, 1996). Whereas we could not detect expression in embryos, *lacZ* expression was observed in ovaries, specifically in follicle cells at an antero-dorsal position, between stages 10 and 14 (Fig. 2B and Table 2). The $l(1)_T$PG31 is an insertion in the *troponin I* (*WupA*) gene, one of several strains which allow to follow the formation of the embryonic somatic musculature (Fig. 2C and Table 2).

During the course of this screen, we employed two distinct and a priori interchangeable transposons as mutagenic elements. We have nonetheless noted certain differences between the PL and PG elements, which we list here. First, significant differences in insertion site selection by the transposase as a function of the element employed may exist. This was most notable for the *inx2* locus, where we obtained a single PL insertion (1/110) but 12 PG insertions (12/169). Second, a much stronger embryonic expression of *lacZ* was observed in $l(1)_T$PG strains crossed to UAS-lacZ than in $l(1)_T$PL strains, but this expression was temporally delayed. Two drawbacks specific to the $l(1)_T$PG lines deserve to be noted: (i) a background expression in embryonic salivary glands from stages 12 to 16 and (ii) a mosaic expression in many tissues, especially in ovarian tissues. Nevertheless, our mutagenesis screen yielded Gal4-expressing lines with very specific expression which can be used to

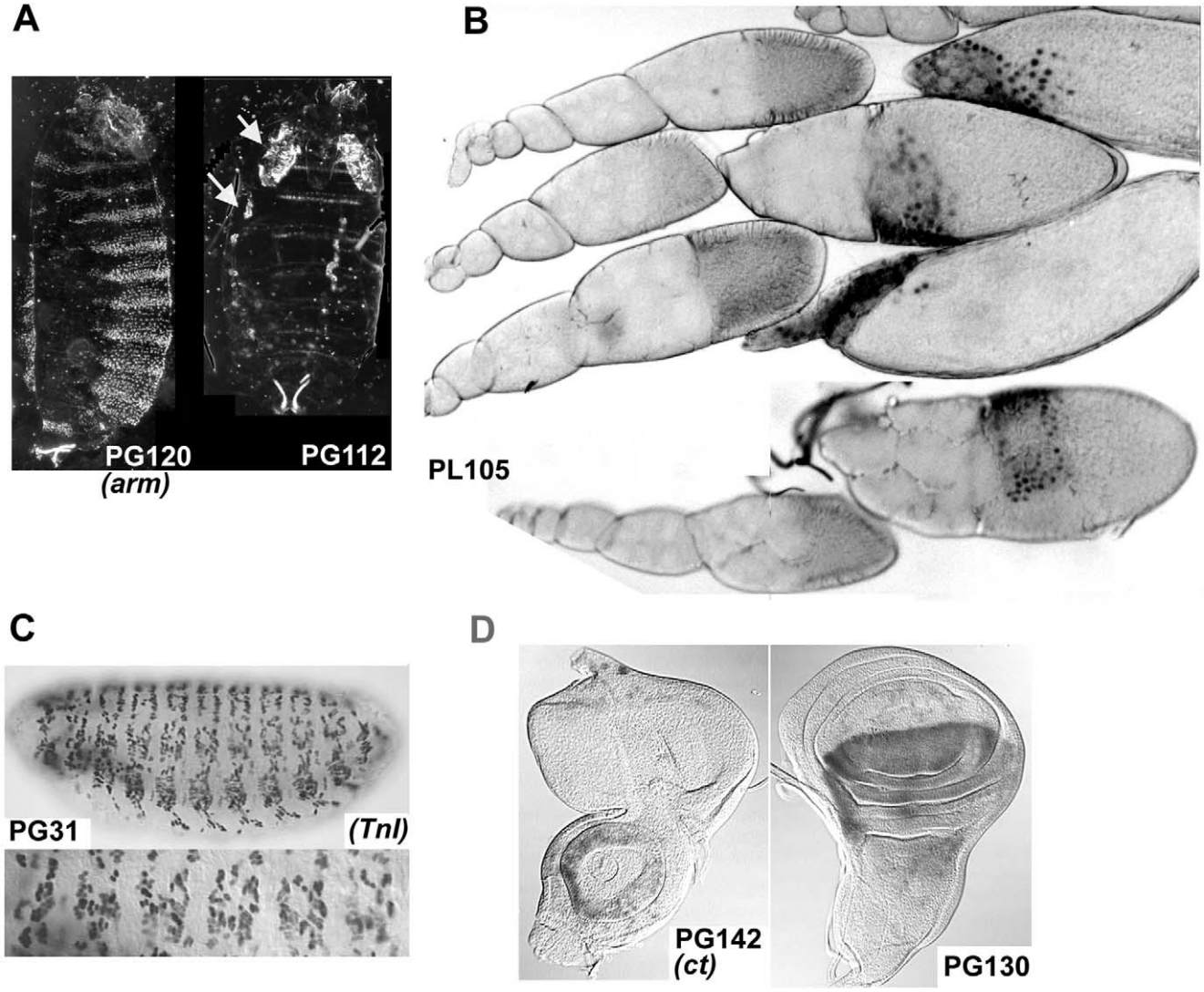

Fig. 2. (A) Dark field photographs of cuticle preparations of late mutant embryos. The white arrows point to melanotic patches found in salivary glands and the dorsal branch of the tracheal system. (B) Dissected ovaries of $l(1)_T$ PL105; *lacZ* expression is specifically observed in antero-dorsal follicle cells, from oogenic stages 10 to 14. (C) *lacZ* expression somatic muscle precursor cells in $l(1)_T$ PG31 embryos. (D) Dissected third instar larval imaginal discs. Left: $l(1)_T$ PG142 eye-antennal disc; *lacZ* expression from this insertion in the *cut* locus is primarily detected in the ring of antennal segment A2. Low-level expression is also detected in the posterior eye. Right: $l(1)_T$ PG130 wing disc; *lacZ* expression is restricted to the dorsal cells. In all panels, the identity of each mutant is indicated by the number of the $l(1)_T$ line and the gene designation, when this information is known.

Table 2
Expression data and phenotypes for the $l(1)_T$ strains[a]

l(1)T line	CG	Embryonic expression	LIII expression	Ovarian expression
PG1	1938	EP; AMS		FC9–14; SC; NC1–8
PG2	1405	EP; NGR		DA, aeropyle
PG3	6769			DA
PG4	4590	EP; CNS; S scattered cells		FC8–14; SC
PG5	14,226	CNS; SPG; AMS		
PG6	4147			FC9–14
PG7	9224	NGR**		
PG8	12,701	EP**		
PG9	3587	PNS; CNS		
PG10	6998	PNS;CNS11-16		
PG11	14,792	CNS; subset of PNS		
PG12	11491			FC7–14, SC
PG13	4281	CNS	EP, FB, TT, GT, CNS, all IGD	
PG14	14,226	APO; CNS		Posterior FC9; tunica
PG15	1405			NC7-8
PG16	9224	NGR**		FC7–14
PG18	9653	NGR**		
PG21	?	MSC	EP, CNS	
PG22	15926		EP, FB, TT, GT, CNS, IGD	
PG23	1751	Hypopharyngeal lobe	EP, FB, TT, GT, CNS, IGD	FC9–14
PG25	1494	MSC (subset); APO	EP, FB, TT, GT, CNS, IGD	FC4–14; BC; SC
PG26	10701	EP; subset of denticle cells; pharynx		FC6–14; SC
PG27	14430	MS10; VML; EP17		FC7–14; SC
PG28	4027	EC; AMS		FC; SC
PG29	7622		CNS	
PG30	1771	CNS14-15, subset; AMGP		FC1–14; NC1-10; stalk; tunica
PG31	7178	MSC14-15	EP, FB, TT, GT, CNS, IGD	
PG32	15926	EP; APO		
PG33	4122	HE; CNS15-17 (subset); AP		
PG34	4147		EP, FB, TT, GT, CNS, IGD	
PG35	14226			FC2-10B; DA; BC; SC; NC11-12
PG37	12,212	PNS**	EP, TT, GT, CNS, IGD	FC8–14; SC; centripetal cells
PG38	17131	EP14-16		
PG39	4147			FC9–12
PG43	7502	EP13-15; target of homeotic gene regulation?		
PG44	2543		EP, FB, TT, GT, CNS, IGD	
PG45	10798	lateral EC		FC8–14;SC
PG46	4590			FC9–14; SC; G1
PG47	4122			BC
PG48	2025		EP, TT, GT, CNS, IGD	SC
PG49	16983		EP, TT, GT, CNS, IGD (not legs)	
PG51	8465	SM; MSC14-16; CNS16	EP, FB, TT, GT, CNS, IGD	FC10–14; NC7-8
PG53	3001	SMS	EP, FB, TT, GT, CNS, IGD	FC9–10; SC
PG55	4590	EP; HG; MT		FC9–14; SC
PG58	11,172	PMG, AMGP; SG; AP		NC869, 12; oocyte9-12; G1
PG59	9012	MSC	EP, FB, TT, GT, CNS, all IGD	FC9–14; SC
PG60	1405	ubiquitous**		FC7–10; BC
PG 61	9214	CNS specific neurons	EP, TT, CNS, T1D and EAD (weak)	
PG62	11190	PNS: PMG	EP, FB, TT, GT, CNS, most IGD	FC9–14
PG63*	9151	CNS; PNS13-15 (subset)	VNC	DA
PG64	9650	SM, MSC12-15	TT, GT, CNS, most IGD	
PG65	4147	AMGP?		FC9–14; DA
PG66	3917			FC10–12
PG67*	10701	EP (denticle cells); GT: OE, HG**	EP, FB, TT, GT, CNS, all IGD	FC9–14; SC
PG68	7981		CNS	
PG69	2849	EP		FC9–14; SC
PG70	2446	PNS; CNS 12-16**	EP, FB, TT, GT, CNS, all IGD	FC9–14; SC

Table 2 (*continued*)

l(1)T line	CG	Embryonic expression	LIII expression	Ovarian expression
PG73	4590			FC9–14; SC
PG75	9650	MSC12-15: CNS	EP, TT, GT, CNS, most IGD	
PG76	3415/9910		EP, GT, CNS	
PG79	1560	TT, TP**	no expression detected	FC11–12
PG80	1521	**	EP, FB, TT, GT, CNS, all IGD	FC8–12
PG81	4590	EC: NGR		
PG83	11,491			FC9–14; NC8-10
PG84	4147			FC9–14
PG85	12202			FC14
PG87	11485	EP		FC9–14; NC6-8
PG89	2849	CNS; AMS; EP		FC9–14
PG90	1405	CNS;MSC (scattered cells); ICP		NC5-9; G1,2
PG91	9653	EC: NGR		
PG95	3981	MSC12-15: EP16	EP, FB, TT, GT, CNS, all IGD	FC9–14; SC; NC7-8
PG96	5870			oocyte 8-10
PG97	14772	MSC13-16	EP, FB, TT, GT, CNS, most IGD	FC9–14; tunica
PG99	9533			FC14; tunica; trachea
PG103	6606	EP		
PG104	4590	EP; ES; MP; HG		
PG105	9650	Mutant phenotype		
PG106	16944			FC7–10
PG107	15194		CNS	
PG108	1405			FC8–14
PG111	11190			FC; SC; tunica
PG112	1521	PNS: SO; CNS**		FC9–14
PG113	1405			FC14; NC5-10
PG115	7113			FC8–14; SC
PG116	9907			DA; NC8
PG117	14619			FC12–14; NC8; tunica
PG118	3719			DA; aeropyle; tunica
PG119	4147	**		FC9–14
PG120	11579	**	FB, GT, CNS	
PG121	9653	NGR; mutant phenotype		
PG122	7981			FC9–14; tunica
PG124	1554		GT, IGD	
PG125	14,226	EP11-16; segmental pattern; APO	EP, TT, GT, BR, IGD	
PG126	15784		EP, FB, TT, GT, CNS, IGD	FC6–14; SC
PG127	10777	EP: OE14-16	EP, FB, TT, CNS, IGD	FC6–14; SC
PG128	11,172	AMS11-14;; AMGP	EP, FB, GT, CNS	
PG129	3861		EP, FB, TT, GT, CNS	
PG130			BR, WGD, HD, EAD	FC14
PG131	5870	AMS11-14; EP10-17	EP, FB, TT, GT, CNS, IGD	FC9–14; SC; tunica
PG132	1903	GT10-16	TT, GT, BR, IGD	
PG133		EP; AMS	EP, FB, TT, GT, CNS, IGD	
PG134	6998	EP11-14; AMS11-14		
PG136	9533		EP, FB, TT, GT, CNS, WGD, HD	
PG137	15047		EP, TT, GT, CNS, WGD, HD, EAD	FC14; tunica; trachea
PG139	6998		BR	DA
PG140	12233	EP; VML scattered cells**	EP, FB, TT, GT, CNS	FC8–14; SC
PG141	16944		FB, GT, CNS	FC8–14
PG142	11387	CNS; GLC 14-17;**	EP, CNS, IGD	
PG143	12238	MSC15-17	EP, FB, GT, CNS	
PG144	4590		EP, TT, CNS, IGD	
PG145	6998		EP, BR	
PG147	14430		EP, FB, TT, GT, CNS, IGD	
PG148	10992	MSC9-12; EP13-17; VML10-12**	EP, GT, CNS	
PG149	3001	MSC13-16	EP, FB, TT, GT, CNS, IGD	
PG150	7981			FC1–14; SC; centripetal cells

Table 2 (*continued*)

l(1)T line	CG	Embryonic expression	LIII expression	Ovarian expression
PG151	1660			DA
PG153	3981	Hypopharyngeal lobe 10-13; MSC13-16; ES14-15	EP, FB, TT, GT, CNS, EAD	FC9–14
PG154	16944		EP, FB, GT, CNS	
PG155	2079	NGR9-12, EPventral12-16; VML	EP, TT, CNS	FC14
PG156	9906	MSC scattered	EP, CNS, most IGD	
PG157	9533	MSC15-16 ES, HG15-16	EP, TT, GT, CNS, WGD (weak)	
PG159	4262		EP, FB, TT, GT, CNS, WGD, HD, EAD	
PG160	4590		EP, TT, CNS, all IGD	
PG161	12359	Mutant phenotype		DA
PG162*	7065	EP dorso lateral10-15; AMS10-13	EP, FB, TT, GT, CNS, most IGD	FC1–14; SC; centripetal cells
PG165	2257		EP, GT, CNS, T1D (weak)	DA
PG166	7981	**	EP, GT, TT, CNS, WGD, HAD, T3D, LBD	operculum, DA
PG167	3001	MSC 13-16	EP, FB, TT, GT, CNS, all IGD	
PG169	13014	CNS15-16; PNS15-16	EP, GT, TT, CNS, EAD (weak)	FC13–14
PL4	2206		GT: PV, PMG; SG, B, FB	NC9-10B
PL5	9413		OE; GT: PV, MG, MT; SG; BR	
PL7		EP (ventral); GT: PH, AMG15 (anterior border)**		
PL9		APO; AMC		
PL11	4147	SM, MSC	PV, MG, MT, SG, B	G2; FC10–14; G2; NC7-9
PL12	6606			FC14; DA
PL13	6998	RS: TT, PS; GT: AMG, BC	PV, MG, MT; SG; MS: FB, RG, GO;	G2-3; FC9–14; SC; CC; NC8-10B
PL15	11937	PNS (subset): MD	RG	
*PL17**	1422	CNS: PNR , VML	CNS (weak)	
PL18*	1740		BR	NC9-10A
PL19	4590	CNS: VML9-10; RS: TP; EP14-17 ; MT	PV, MG,; SG; BR; IGD	
PL20	7952	procephalic lobe; gnathal and A5-A7 segments**		
PL23	1405		PV, MG, SG	FC14; NC9
PL24	3936	**	BR; IGD; histoblasts	FC6–14; SC; CC; BC
PL25	18319			BC; stalk cells; posterior cell
PL26	11190		BR; SG	FC; NC9-10
PL28	4068	PNS15: AMC; ES		FC8–14, DA; D-V gradient at stage 10A
PL29	7307	CNS; PNS (labral segment)	IGD; CNS	centripetal and squamous FC
PL32	1725	EP; AMS	BR	
PL33	14792	PNS: AMC; GT	IGD; CNS	FC9–10; NC7-10B
PL35	10798		PV; CNS	
PL38	3665	RS11-14 ; TP ,PS; PNS: AMC	MT; SG; IGD; GO; BR (optic lobes)	G1-3; FC-7; BC
PL40	7893	CNS: VML; PNS (few cells)		
PL41*	14226	CNS; PNS: AMC	PV; CNS	
PL42	9906	PNS: SO; CNS	CNS	
PL43	9606		CNS	
PL44	8497	dorsal pouch		
PL45	9725 /9732	EP15; GT: HG; CNS	EP; SG; CNS	
PL48	3525		MG; SG; IGD; CNS (optic lobes)	NC9-10
PL49	2059/1677	CNS (few cells)	optic lobes	G2-3; FC
PL50	11190	ES; CNS (weak); PNS: AMC	PV; CNS	FC; dorsal 10B; G1
PL51	10798	SM (weak), DPM; VM; AMG, PMG, HG	PV; SG; IGD; BR (optic lobes)	FC
PL53	2206			FC8-10B; centripetal cells
PL54	10701	GT: PV, cells bordering the PH; PNS: AMC		FC4–7, 14
PL55#1#2	6606	RS: PS; SG; EC; CNS VML	EP; CNS; IGD	NC6-10B

Table 2 (*continued*)

l(1)T line	CG	Embryonic expression	LIII expression	Ovarian expression
PL56#1#2	2849	PNS: SO, AMC	MG; PV	FC10–14, DV gradient; SC; CC; DA
PL57	1903	CNS (weak)	SG; BR	NC9-11
PL60#1 #2	15047	GT14: BC; CNS15	SG phenotype; IGD phenotype; BR	
PL62	7434		BR; RG	
PL64*	9533	GT: ES, MP, HG; MSC: precursors (few)	PH; TT; SG; BR	FC13–14; DA
PL66	11 485		BR	
PL67	12212	PNS; AMS**		
PL68	4147	SG; GT: AMGP; MS: FB		FC9–14; NC9-10
PL69	4349			DA; NC9-10
PL71	11190	GT: ES		
PL72	6606	CNS (weak), VML; GT: PH		FC12–14; CC10B; DA
PL73	4147	SG; GT: AMGP	SG; MG	NC6-10; G2
PL74	9224	NGR; CNS: VML; EP (late segmental pattern)**	histoblasts; IGD;	
PL75*	7326	PNS: SNS; RS: TP	BR; IGD	FC10–14; CC; NC8-10
PL76	3774	AMGP		
PL77	9725/9732	PNS; CNS (weak); PV	SG	FC14; DA
PL78	16902	GT: PH, HG**		FC14; DA
PL79	14792		CNS	FC10–14; NC8_10
PL82	1405	CNS; PNS: AMC	PV; MG; SG; FB	
PL83	12225			NC8-10
PL84	3665	RS11-14 ; TP, PS ; GT: HG, PH; MS: RG	MT; IGD	FC-8; BC posterior cell
PL86	11491		OE; SG	
PL87	7178	MSC 15-16		
PL88	1740	CNS: SPG; PNS: AMC; GO		
PL91	3595		MG; PV; CNS	FC4-10 anterodorsal; NC10
PL92	12701	IGD: PD, WGD, HD; CNS16 (weak)**	CNS; IGD	
PL93	14792			FC7–14; NC3-11
PL94	17841		RG; SG	
PL97	1405		CNS; SG; FB; PMG	FC10–14; antero-dorsal; NC7-10A
PL100	14226		IGD	Posterior pole at early stages; BC
PL101	12235	AMS14		
PL104	15365	MS: HE	IGD	DA; aeropyle
PL105	5530		BR	FC7–8; anterodorsal10-12; DA
PL106	10701	EP15; AP	EP; MG; IGD	FC6–8; 14
PL107	6824	EP (denticle cells)**	EP; BR	FC1–7
PL109	7981		MG; SG; FB; BR	FC5–14; SC

[a] All stocks from the $l(1)_T$ collection showing expression of the reporter gene in either ovaries, embryos, third instar larvae or a combination of, are listed by strain number. An asterisk in the first column indicates strains for which a significant proportion of male survivors with red eyes is recovered after each cross. A double asterisk in the embryonic expression column indicates mutant strains displaying embryonic lethality with recognizable cuticular defects. Mutant phenotype indicates strains where LacZ staining reveals tissue-specific defects. The predicted CG mutated is recalled in the second column. The expression patterns are described using a simple key-word nomenclature corresponding to the different tissues examined (see list below), sometimes followed by the specific stages when expression is observed. Embryonic and larval expression: SM, somatic mesoderm; MSC, somatic muscles; VM, visceral mesoderm; DV, dorsal vessel; CA, cardiac cells; RG, ring gland; FB, fat body; HE, hemocytes; GT, gut; PH, pharynx; ES, esophagus; PV, proventriculus; AMG, anterior midgut; PMG, posterior midgut; HG, hindgut; MP, Malpighian tubules; ICP, interstitial cell progenitors; AMGP, adult midgut progenitors; BC, basophilic cells; EP, epidermis; NGR, neurogenic region; OE, oenocytes; IGD, imaginal discs; PD, prothoracic disc; WGD, wing disc; HD, haltere disc; AP, anal pads; APO, apodemes; SG, salivary gland; AMS, amnioserosa; PNS, peripheral nervous system; SO, sensory organs; AMC, antennomaxillary complex; MD, multi-dentritic neurons; SNS, stomato nervous system; CNS, central nervous system; VNC, ventral nerve cord; GLC, glial cells; PNR, procephalic neuroectoderm; BR, brain; MEC, mesectoderm; VML, ventral midline; TT, tracheal tree; TP, tracheal placodes; PS, posterior spiracles; GO, gonads. Ovarian expression: NC, nurse cells; FC, follicle cells; DA, dorsal appendages; BC, border cells; G, germarium; DV, dorsoventral.

trace a variety of cell lineages and modify their program of gene expression, including in imaginal discs, as illustrated in Fig. 2D. $l(1)_T$PG142 is an insertion in the *cut* (*ct*) gene which encodes a homeodomain transcription factor (Wieschaus et al., 1984; Blochlinger et al., 1988). Strong $l(1)_T$PG130 expression in the wing disk, especially the

prospective notum and wing blade, is restricted to dorsal cells. $l(1)_T$PG130 is inserted within the long terminal repeat of a yoyo transposon whose chromosomal location has not been determined; of four PG lethal insertions located within another transposable element, only $l(1)_T$PG130 has been kept in our collection because of this expression. The lethality associated with PG insertions may also provide a direct assay for the function of the putative mutated gene, using UAS-cDNA constructs in phenotypic rescue experiments. This assay has been successfully used to verify and phenotypically characterize mutations in the *moesin* gene (Cedric Polesello and François Payre, personal communication).

3. Materials and methods

3.1. Dysgenic crosses and selection of X-linked lethal P-insertions

Flies were grown on standard corn meal/agar medium at 25°C. Two different P-element vectors were used: PlacW (Bier et al., 1989) and PGawB (Brand and Perrimon, 1993) (PL and PG strains). Before mutagenesis, the targeted X-chromosome was isogenized and selected for good viability and fertility in both males and females (*y w* (iso)). Cross 0: female *y w*(iso); CyO, *Cy* PlacW/*Sp* or *y w* (iso); CyO, *Cy* PGawB/*Sp* flies were crossed with *y w* (iso); TMS, *Sb* Δ2–3/Dr^{Mio} males. Cross 1: two dysgenic *y w* (iso); CyO, *Cy* P/+; TMS, *Sb* Δ2–3/+ female progeny (variegated eye color) were crossed in single vials with two males carrying the marked X balancer FM7, *y w B*. These females carried two isogenized X chromosomes, the mobile P element marked with mini-*white* and the transposase source. The progeny of each cross was examined daily during 5 days for [Cy^+ Sb^+ mini-*white*$^+$ *B*] virgin females harboring a transposed P element but lacking the starting P and transposase sources carried by the CyO, *Cy* and TMS, *Sb* chromosomes, and the *B* marker from the FM7 chromosome. To avoid mutation clustering, Cross 1 tubes were discarded once a single mini-w^+, *B* virgin candidate female was recovered. Cross 2: candidate *y w* (iso) P?/FM7, *y w B* females were re-crossed with FM7, *y w B* males. The male progeny of this cross falls into three main classes: (i) new autosomal insertions produce B^+ males both with and without the mini-w^+ marker; (ii) for viable X insertions, all B^+ males harbor the mini-w^+ marker; (iii) for lethal insertions in essential X loci, all males are FM7, *y w B*. In addition to fully lethal insertions, we also kept lines with clearly reduced male viability (few B^+ males compared to the FM7 reference chromosome; see Table 1). The mutant strains of our collection were designated $l(1)_T$PL(1–110) or $l(1)_T$PG(1–169), for PlacW and PGawB inserts, respectively. To ensure that newly induced recessive lethality was attributable to the new P insertion, the P elements in a number of lines were re-mobilized by the Δ2–3 transposase, including at least one $l(1)_T$PL or $l(1)_T$PG line for each gene for which no mutation had previously been described. Isolation of viable white-eyed males among the progeny indicated that the P insertion was indeed the cause of lethality. Among 111 lines tested, 91 were successfully mobilized to generate fully viable males. In the other strains, we were unable to rescue lethality. Since these strains included both PL and PG lines, and were randomly distributed in the temporal order of their isolation, we conclude that most lethal mutations can be attributed to the P insertion.

3.2. Enhancer trap patterns

Each individual strain was stained for enhancer trap patterns in embryos and ovaries, and (in most cases) the imaginal discs of third instar larvae, as indicated in Table 2. Embryonic and larval *lacZ* expression was detected using an antibody directed against β-galactosidase (Promega; 1/1000 dilution). Visual inspection of the LacZ patterns also served to identify major patterning and/or tissue differentiation defects. Alternatively, larval and adult patterns were detected by histochemical detection of β-galactosidase activity according to Yanicostas et al. (1995). In cases of embryonic lethal or semi-lethal mutations, cuticle preparations of unhatched embryos were prepared as indicated in Wieschaus et al. (1994).

3.3. Flanking sequence determination

Localization of inserts was accomplished via determination of genomic sequence flanking the insert and comparison with the whole genome sequence. Sequences flanking all P-element insertions in our collection were determined from either one or both ends by one of two methods, the plasmid rescue procedure (Pirrotta, 1986) and inverse polymerase chain reaction (PCR) (http://www.fruitfly.org/methods/).

3.4. Association with ESTs and CG numbers

Identification of nearby CGs was done by comparing the sequence flanking each P-insertion with the most recently annotated genomic DNA, using the BLAST software at BDGP. For most lines, the insertion was located within or close to the 5′-region of a predicted gene, allowing a straightforward link to a specific CG. In other, less direct cases, we searched for the possible existence of introns not previously annotated introns by looking at the structure of available ESTs sequenced at both ends and/or potential consensus acceptor splice sites located upstream of the proposed translation initiator ATG. In the cases of insertions located between two CGs transcribed from the same DNA strand, we considered that the CG located downstream of the P-insertion site was most likely to be the mutated gene. All details on the annotation of the $l(1)_T$ collection are available upon request to bourbon@cict.fr.

3.5. Stock distribution

At present, lines from the primary $l(1)_T$ collection are maintained at the Centre de Biologie du Développement in Toulouse and available upon request to cbd_dir@cict.fr. Strains corresponding to genes presently studied in our laboratories are indicated in bold in Table 1. Return information from further study of any of the $l(1)_T$ stocks is very much welcome and should be forwarded for display on our website using the request procedure.

Acknowledgements

We especially wish to thank Marie Louise Dumont, Nathalie Firmin, Agnes Lepage, Marie-Josephe Guinaudy, Jacqueline Becker, Nelly Léger, for excellent technical assistance and Bruno Savelli for his invaluable help in creating our $l(1)_T$ website. We also thank Michèle Crozatier, François Payre, and Sandrine Peyrefitte for sharing unpublished results. This work was supported by the Centre National de la Recherche Scientifique (Programme Genome).

References

Adams, M.D., Celniker, S.G., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., 1989. Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. Genes Dev. 3, 1273–1287.

Blochlinger, K., Bodmer, R., Jack, J., Jan, L.Y., Jan, Y.N., 1988. Primary structure and expression of a product from *cut*, a locus involved in specifying sensory organ identity in *Drosophila*. Nature 333, 629–635.

Brand, A., Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401–405.

Casal, J., Leptin, M., 1996. Identification of novel genes in *Drosophila* reveals the complex regulation of early gene activity in the mesoderm. Proc. Natl Acad. Sci. USA 93, 10327–10332.

Glise, B., Bourbon, H.M., Noselli, S., 1995. *hemipterous* encodes a novel *Drosophila* MAP kinase kinase, required for epithelial cell sheet movement. Cell 83, 451–461.

Glodstein, L.S., Gunawardena, K., 2000. Flying through the *Drosophila* cytoskeletal genome. J. Cell Biol. 150, F63–F68.

Gonzy-Treboul, G., Lepesant, J.A., Deutsch, J., 1995. Enhancer-trap targeting at the broad-complex locus of *Drosophila melanogaster*. Genes Dev. 9, 1137–1248.

Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S., Rushlow, C., 1999. The *Drosophila* gene *brinker* reveals a novel mechanism of Dpp target gene regulation. Cell 19, 563–573.

Lasko, P., 2000. The *Drosophila melanogaster* genome: translation factors and RNA binding protiens. J. Cell Biol. 150, F51–F56.

Levin, L.R., Han, P.L., Hwang, P.M., Feinstein, P.G., Davis, R.L., Reed, R.R., 1992. The *Drosophila* learning and memory gene *rutabaga* encodes a Ca^{2+}/Calmodulin-responsive adenylyl cyclase. Cell 68, 479–489.

Mevel-Ninio, M., Terracol, R., Salles, C., Vincent, A., Payre, F., 1995. *ovo*, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. Mech. Dev. 49, 83–95.

Miklos, G.L., Rubin, G.M., 1996. The role of the genome project in determining gene function: insights from model organisms. Cell 86, 521–529.

Mount, S.M., Salz, H., 2000. Pre-messenger RNA processing factors in the *Drosophila* genome. J. Cell Biol. 150, F37–F43.

Payre, F., Vincent, A., Carreno, S., 1999. *ovo/svb* integrates Wingless and DER pathways to control epidermis differentiation. Nature 400, 271–275.

Peyrefitte, S., Kahn, D., Haenlin, M., 2001. New members of the *Drosophila Myc* transcription factor subfamily revealed by a genome-wide examination for basic helix–loop–helix genes. Mech. Dev. 104, 99–104.

Pirrotta, V., 1986. Cloning *Drosophila* genes. In: Roberts, D.B. (Ed.). *Drosophila*: A Practical Approach. IRL Press, Oxford, pp. 83–110.

Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G.M., Weigmann, K., Milan, M., Benes, V., Ansorge, W., Cohen, S.M., 1998. Systematic gain-of-function genetics in *Drosophila*. Development 125, 1049–1057.

Rubin, G.M., Lewis, E.B., 2000. A brief history of *Drosophila's* contributions to genome research. Science 287, 2216–2218.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor, Miklos, G.L., Nelson, C.R., et al., 2000. Comparative genomics of the eukaryotes. Science 287, 2204–2215.

Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S., Rubin, G.M., 1999. The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosphila* genes. Genetics 153, 135–177.

Vincent, A., Blankenship, J., Wieschaus, E., 1997. Integration of the head and trunk patterning systems controls cephalic furrow formation in *Drosophila*. Development 124, 3747–3754.

Wieschaus, E., Nüsslein-Volhard, C., Jürgens, G., 1984. Mutations affecting the pattern of the larval cuticle on the X-chromosome and fourth chromosome. Wilhem Roux's Arch. 193, 296–307.

Yanicostas, C., Ferrer, P., Vincent, A., Lepesant, J.A., 1995. Separate cis-regulatory sequences control expression of *serendipity beta* and *janus A*, two immediately adjacent *Drosophila* genes. Mol. Gen. Genet. 246, 549–560.